MDPI

Article

Generalizing the Balance Heuristic Estimator in Multiple Importance Sampling

Mateu Sbert [1,*] and **Víctor Elvira** [2,3]

1 Informatics and Applications Institute, Girona University, 17003 Girona, Spain
2 School of Mathematics, University of Edinburgh, Edinburgh EH8 9YL, UK; victor.elvira@ed.ac.uk
3 The Alan Turing Institute, London NW1 2DB, UK
* Correspondence: mateu@ima.udg.edu

Abstract: In this paper, we propose a novel and generic family of multiple importance sampling estimators. We first revisit the celebrated balance heuristic estimator, a widely used Monte Carlo technique for the approximation of intractable integrals. Then, we establish a generalized framework for the combination of samples simulated from multiple proposals. Our approach is based on considering as free parameters both the sampling rates and the combination coefficients, which are the same in the balance heuristics estimator. Thus our novel framework contains the balance heuristic as a particular case. We study the optimal choice of the free parameters in such a way that the variance of the resulting estimator is minimized. A theoretical variance study shows the optimal solution is always better than the balance heuristic estimator (except in degenerate cases where both are the same). We also give sufficient conditions on the parameter values for the new generalized estimator to be better than the balance heuristic estimator, and one necessary and sufficient condition related to χ^2 divergence. Using five numerical examples, we first show the gap in the efficiency of both new and classical balance heuristic estimators, for equal sampling and for several state of the art sampling rates. Then, for these five examples, we find the variances for some notable selection of parameters showing that, for the important case of equal count of samples, our new estimator with an optimal selection of parameters outperforms the classical balance heuristic. Finally, new heuristics are introduced that exploit the theoretical findings.

Keywords: Monte Carlo; importance sampling; balance heuristic; variance reduction; chi-square divergence; Kullback–Leibler divergence; cross entropy

---

1. Introduction

Multiple importance sampling (MIS) is a Monte Carlo technique widely used in the literature of signal processing, computational statistics, and computer graphics for approximating complicated integrals. In its basic configuration, it works by drawing random samples from several proposal distributions (also called techniques) and weighting them appropriately in such a way that an estimator built with the pairs of weighted samples is consistent. Since the publication of [1], the celebrated *balance heuristic* estimator has been extensively used in the Monte Carlo literature, with an unprecedented success in the computer graphics industry (Eric Veach has been awarded with several prizes because of his contributions in the MIS literature, where the balance heuristic is arguably the most relevant one). In the balance heuristic method, different samples are simulated from each proposal and the traditional IS weight is assigned to each of them. Unlike the standard IS estimator, all the weighted samples are combined with an extra weighting, in such a way the resulting estimator typically shows a reduced variance. Its superiority in terms of variance with regard to other traditional combination schemes has been recently shown in [2], where a framework is established for sampling and weighting in MIS under equal number of samples per technique. The balance heuristic, also called the *deterministic mixture* [3], has been widely used in the literature of MIS. Further efficient variance reduction techniques

are proposed in [4–6] also in the context of MIS, still with equal counts from each technique. Provably better estimators [7] and heuristically better ones [8,9] have been presented that use a different count of samples than equal count for all techniques. In [10], it has been shown the relationship of a better count of samples with generalized weighted means. The balance heuristic is also present in most of successful adaptive IS (AIS) methods, see [11–15], in particular in the case where all techniques are used to simulate the same number of samples. Recently, MIS has returned to the main focus in computer graphics in [16], where it is shown that allowing weights to be negative, the reduction over balance heuristics of the resulting MIS estimator can be higher than the one predicted by Veach bounds [17]; in [18], where one of the sampling techniques is optimized to decrease the overall variance of the resulting MIS estimator; in [19], where the weights are made proportional to the quotients of second moments divided by the variances of the independent techniques; in [20], where MIS is generalized to uncountably infinite sets of techniques.

Interestingly, the balance heuristic has two properties in the assigned weights. First, all techniques appear at the denominator of the weight of a specific technique. Second, they appear in a form of a mixture, with coefficients proportional to the number of samples simulated from each technique. In this paper, we relax this constraint providing a generalized weighting/combining family of estimators that has the balance heuristic as a particular case. First, we show that it is possible to use a specific set of coefficients to decide the amount of samples per technique, and a different set of coefficients to be applied as the importance weight. Second, we study four different cases fixing some of these coefficients (sampling and/or weighting), and we give the optimal solution for the rest of coefficients in such a way the variance of the MIS estimator is minimized. Note that, the novel estimator always outperforms the balance heuristic under the optimal choice of those coefficients. In five numerical examples we show that, under an adequate choice of parameters, the novel estimator outperforms the celebrated balance heuristic.

The rest of the paper is structured as follows. Section 2 revisits the balance heuristic estimator. In Section 3, we propose the new family of estimators that generalizes the balance heuristic. We address five cases of special interest, depending on which parameters are free and the number of samples simulated from each technique. We then, in Section 4, discuss the different singular points in the variances of the estimators considered. In Section 5, we give a necessary and sufficient condition, in terms of χ^2 divergence (and its approximation as a Kullback–Leibler divergence), for the variance of the new estimator to be better than the one of balance heuristics. Finally, we conclude the paper with five numerical examples in Section 6, propose some heuristics, and give some conclusions in Section 7.

2. Balance Heuristic Estimator

The goal in IS is usually the estimation of the value of integral $\mu = \int f(\mathbf{x})d\mathbf{x}$. In MIS, n_i samples, $\{\mathbf{X}_{i,j}\}_{j=1}^{n_i}$, are simulated from a set of available probability density functions (pdfs), $\{p_i\}_{i=1}^{n}$. The MIS estimator introduced by Veach and Guibas [1] is given by

$$Z = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} w_i(\mathbf{X}_{i,j}) \frac{f(\mathbf{X}_{i,j})}{p_i(\mathbf{X}_{i,j})}, \tag{1}$$

where $w_i(\mathbf{x})$ is a weight function associated to the i-th proposal that fulfills both following conditions. First, the weights must sum up to one in all points of the domain where the value of the function is different from zero, i.e., $\sum_{i=1}^{n} w_i(\mathbf{x}) = 1$, $\forall x$ where $f(\mathbf{x}) \neq 0$, Second, for all x where $p_i(\mathbf{x}) = 0$, then $w_i(\mathbf{x}) = 0$. In this paper, we consider only weighting functions that yield Z unbiased.

The *balance heuristic* estimator is a particular case of Equation (1) where the weight function is given by

$$w_i(\mathbf{x}) = \frac{n_i p_i(\mathbf{x})}{\sum_{k=1}^{n} n_k p_k(\mathbf{x})}, \tag{2}$$

which can be written too as

$$w_i(\mathbf{x}) = \frac{\alpha_i p_i(\mathbf{x})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})}, \tag{3}$$

where $n_i = \alpha_i N$. In this case, the estimator in Equation (1) becomes the balance heuristic or deterministic mixture estimator given by

$$F = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{\alpha_i f(\mathbf{X}_{i,j})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{X}_{i,j})} \tag{4}$$

$$= \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{f(\mathbf{X}_{i,j})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{X}_{i,j})}. \tag{5}$$

2.1. Interpretation of F and General Notation of the Paper

Note that F is also called *deterministic mixture scheme* [3] or *multi-sample MIS estimator* [7,17], as opposed to the case where all independent and identically distributed (i.i.d.) samples are from the mixture $\psi_\alpha = \sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})$. The latter alternative is called the *randomized balance heuristic* or *one-sample MIS estimator*, and the estimator is denoted by $\mathcal{F}$. This alternative is a subtle variation of F that simply modifies the sampling by simulating the N total samples i.i.d. from the mixture ψ_α (instead of deterministically choosing the number of samples per proposal). Then, the estimator is the exact same expression of Equation (4). This randomized version $\mathcal{F}$ allows also for a re-interpretation of the deterministic balance heuristic, F, that can be seen as a mixture sampling (note that the mixture is present in the denominator of all importance weights) with variance reduction using stratified sampling (see Appendix 1 of [2] for a detailed discussion). In the randomized version, n_i is then a random variable with expected value $\alpha_i N$, for all i. In summary, in this paper we refer as *deterministic mixture estimator* the cases where the number of samples per technique is deterministic, e.g., F, and *random mixture estimator* when there is i.i.d. sampling from the mixture (and hence n_i is a random variable), e.g., $\mathcal{F}$. We represent the randomized estimators with calligraphic letters. All estimators, unless the opposite is clearly stated, are deterministic mixture estimators.

Finally, we denote estimators with the superindex[1] when they are versions of a specific estimator but with a number of samples normalized to 1, e.g., F^1 is the normalized version of F. In other words, even if the estimators require that all the numbers of samples per technique are $n_i \in \mathbb{R}$, we use these normalized estimators to denote the variance normalized to 1 sample, which simplifies the comparison across estimators (for N total samples, the variance of the estimator would be just the variance of F^1 divided by N).

In Table 1 we show the naming convention used in this paper.

Table 1. Naming convention for the multiple importance sampling estimators in this paper. We drop the superindex[1] from primary estimators when not strictly necessary.

Z	Generic deterministic (multi-sample) MIS estimator
Z^1	Generic deterministic (multi-sample) MIS estimator normalized to one sample
$\mathcal{Z}$	Generic randomized (one-sample) MIS estimator
$\mathcal{Z}^1$	Generic randomized (one-sample) MIS estimator, for number of samples equal to 1
F	Balance heuristic multi-sample MIS estimator
F^1	Balance heuristic multi-sample MIS estimator normalized to one sample
$\mathcal{F}$	Balance heuristic one-sample MIS estimator
$\mathcal{F}^1$	Balance heuristic one-sample MIS estimator, for number of samples equal to 1
G	Generalized balance heuristic multi-sample MIS estimator
G^1	Generalized balance heuristic multi-sample MIS estimator normalized to one sample
$\mathcal{G}$	Generalized balance heuristic one-sample MIS estimator

2.2. Rationale

In Theorems 9.2 and 9.4 of [17], the variances of Z^1 and its randomized version $\mathcal{Z}^1$ are given, i.e., the version where instead of deterministically selecting n_i, all the samples are directly simulated from $\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})$. By subtracting their values we have

$$V[\mathcal{Z}^1] - V[Z^1] = \sum_i \alpha_i \mu'^2_i - \mu^2, \tag{6}$$

where (we use here a notation to be consistent with notation in [7])

$$\mu'_i = \frac{1}{\alpha_i} \int w_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \tag{7}$$

and thus

$$\sum_i \alpha_i \mu'_i = \mu. \tag{8}$$

This result generalizes several particular cases derived in [7] and [2], in the context of a variance analysis of MIS estimators. From Equation (6) (see Appendix A) we have that

$$V[Z^1] \leq V[\mathcal{Z}^1], \tag{9}$$

and equality only happens (apart from the case when both variances $V[Z^1]$, $V[\mathcal{Z}^1]$ are zero) when for all i all μ'_i are equal. One example is given by taking in Equation (7) for all i, $w_i(\mathbf{x}) = w_i$ constant and $\alpha_i = w_i$. We also show in Appendix A that, for the particular case when $\alpha_i = 1/n$ and $f(x) \geq 0$, the upper bound for improvement of deterministic versus non-deterministic estimator is given by

$$V[\mathcal{Z}^1] - V[Z^1] \leq (n-1)\mu^2, \tag{10}$$

where the bound would be approached when there is an index k such that for all $i \neq k$, $\mu'_k >> \mu'_i$. In Theorem 9.4 of [17], it is shown that the optimal weighting functions $w_i(\mathbf{x})$ for $\mathcal{Z}$ (i.e., those that minimize their variance $V[\mathcal{Z}]$), are the balance heuristic weights, Equation (3); therefore, the optimal case is $\mathcal{Z} \equiv \mathcal{F}$, where $\mathcal{F}$ is the random mixture estimator. This is, for any estimator $\mathcal{Z}$, when using the same distribution of samples, also taking into account Equation (9) (see also [7]), it always holds that

$$V[F] \leq V[\mathcal{F}] \leq V[\mathcal{Z}]. \tag{11}$$

Further, in Theorem 9.2 of [17], it is proved that the estimator that optimizes the second moment of Z^1 estimator, this is, $V[Z^1] + \sum_i \alpha_i \mu'^2_i$, is the balance heuristic estimator. Thus, it seems clear that for improvement we have to look for a deterministic estimator, which should be a generalization of balance heuristic mixture estimator F. This is presented in the next section.

2.3. Realistic Applications

2.3.1. Global Illumination

The rendering equation [21] tells us that the radiance $L_{out}(x, \omega_{out})$ exiting point x in direction ω_{out} is given by

$$L_{out}(x, \omega_{out}) = L_e(x, \omega_{out}) + \int_\Omega L_{in}(x, \omega_{in}) \rho(x, \omega_{in}, \omega_{out}) \cos\theta d\omega_{in}, \tag{12}$$

where $L_e(x, \omega_{out})$ is the emitted radiance, $L_{in}(x, \omega_{in})$ is the incoming radiance at x from direction ω_{in}, with the integral extended to the hemisphere Ω above x, θ is the angle of the normal at x with ω_{in}, and ρ is the bidirectional reflectance distribution function, which gives the fraction of luminance incoming from ω_{in} and scattered in the direction ω_{out}. To

approximate the integral in (12), multiple importance sampling is applied with pdfs that sample, respectively, the incoming radiance and the bidirectional reflectance.

2.3.2. Bayesian Inference

Multiple importance sampling (MIS) is often applied in Bayesian inference [22]. In this context, the posterior distribution is usually intractable, and computing its moments analytically is impossible; therefore, approximate methods are required. MIS is particularly well suited for cases when the posterior distribution is multimodal or needs to be approximated by a mixture of densities [2].

Typically, a set of observations, $\mathbf{y} \in \mathbb{R}^{d_y}$, are available. The inference task consists in estimating probabilistically some hidden parameters and/or latent variables $\mathbf{x} \in \mathbb{R}^{d_y}$, which are connected to the observations through a statistical model. The relation between observations and unknown parameters is encoded in the likelihood, $\ell(\mathbf{y}|\mathbf{x})$, and the prior knowledge on $\mathbf{x}$ is described through the prior distribution $p_0(\mathbf{x})$. Through the Bayes' rule, we can obtain the posterior distribution of the unknowns as

$$\widetilde{\pi}(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x}) p_0(\mathbf{x})}{Z(\mathbf{y})}, \tag{13}$$

where $Z(\mathbf{y}) = \int \ell(\mathbf{y}|\mathbf{x}) p_0(\mathbf{x}) d\mathbf{x}$ is the marginal likelihood.

In this scenario, we can be interested in computing a moment $h(\mathbf{x})$ of the posterior, which is an expectation (hence an integral) of the following form:

$$\mu = E_{\widetilde{\pi}}[h(\mathbf{x})] = \int h(\mathbf{x}) \widetilde{\pi}(\mathbf{x}|\mathbf{y}) d\mathbf{x}, \tag{14}$$

where $h : \mathbb{R}^{d_x} \to \mathbb{R}$. In connection with the notation of this paper, $f(\mathbf{x}) = h(\mathbf{x})\widetilde{\pi}(\mathbf{x}|\mathbf{y})$. In the context of Bayesian inference, there are several ways to select the techniques, $\{p_i(x)\}_{i=1}^n$. For instance, they can be deterministically set a priori, e.g., as Laplace approximations at different points [23]. Alternatively, they can be adapted over iterations via iterative stochastic mechanisms [24]. The latter technique is called adaptive importance sampling (AIS) and its literature is vast (see [15] for a recent review).

3. Generalized Multiple Importance Sampling Balance Heuristic Estimator

In classic balance heuristic, both the number of samples $\{n_i = \alpha_i N\}_{i=1}^n$ and the weights of the mixture of pdfs, $\{\alpha_i\}_{i=1}^n$, are given by the same set of parameters. Let us consider now the estimator of Equation (4), where we relax the dependence between the number of samples n_i, and the associated coefficient α_i, i.e., now $n_i = \beta_i N$, $\beta_i > 0$, $\sum_{i=1}^n \beta_i = 1$, where in general $\alpha_i \neq \beta_i$ (otherwise, we recover F). We now define the estimator

$$
\begin{aligned}
G &= \sum_{i=1}^n \frac{\alpha_i}{n_i} \sum_{j=1}^{n_i} \frac{f(\mathbf{X}_{i,j})}{\sum_{k=1}^n \alpha_k p_k(\mathbf{X}_{i,j})} \tag{15} \\
&= \frac{1}{N} \sum_{i=1}^n \frac{\alpha_i}{\beta_i} \sum_{j=1}^{n_i} \frac{f(\mathbf{X}_{i,j})}{\sum_{k=1}^n \alpha_k p_k(\mathbf{X}_{i,j})}. \tag{16}
\end{aligned}
$$

Note that G is a particular case of Z, with weights $w_i = \frac{\alpha_i p_i(\mathbf{x})}{\sum_{k=1}^n \alpha_k p_k(\mathbf{X}_{i,j})}$ in Equation (1). Note also that the balance heuristic F is a particular case of G, i.e., in general we do not impose the restriction of $\alpha_i = \frac{n_i}{N}$. Although the allocation of samples will be different for F and G, we remark that both estimators require the same number of simulations and evaluations. Finally, we note that a close examination of (16) allows us to re-interpret the importance weight associated to each sample. The mixture parametrized with $\{n_i = \alpha_i N\}_{i=1}^n$ appears in its denominator, but since this mixture does not reflect the sampling procedure (the true mixture proposal in sampling is the one parametrized by $\{n_i = \beta_i N\}_{i=1}^n$), the importance

weight of a sample simulated from the technique β_i, is corrected by the factor α_i/β_i in order to account for this mismatch.

Estimator G can be rewritten as $G = \sum_{i=1}^{n} \alpha_i G_i$, where

$$G_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{f(\mathbf{X}_{i,j}))}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{X}_{i,j}))}. \tag{17}$$

Note also that G depends of two sets of parameters, $\{\alpha_i\}_{i=1}^{n}$ and $\{\beta_i\}_{i=1}^{n}$. In the particular case where $\beta_i = \alpha_i, \forall i$, the estimator G becomes F. Let us consider the case with $n_i = 1$. Then,

$$G_i' = \frac{f(\mathbf{x})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})}, \tag{18}$$

with expectation

$$E[G_i'] = \int \frac{f(\mathbf{x}) p_i(\mathbf{x})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})} d\mathbf{x} \equiv \mu_i', \tag{19}$$

and variance

$$\sigma_i'^2 = \int \frac{f^2(\mathbf{x}) p_i(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x} - (\mu_i')^2. \tag{20}$$

Observe that $E[G_i] = E[G_i'] = \mu_i'$, and $V[G_i] = \frac{1}{n_i} V[G_i'] = \frac{1}{n_i} \sigma_i'^2$.

Theorem 1. *For any set of weights $\{\alpha_i\}_{i=1}^{n}$, such as $\sum_{i=1}^{n} \alpha_i = 1$ and any set of weights $\{\beta_i\}_{i=1}^{n}$, such as $\sum_{i=1}^{n} \beta_i = 1$, G is an unbiased estimator of μ.*

Proof. The estimator G is unbiased, since

$$E[G] = \sum_{i=1}^{n} \alpha_i E[G_i] = \sum_{i=1}^{n} \alpha_i \mu_i' = \sum_{i} \alpha_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})} d\mathbf{x} \tag{21}$$

$$= \int \frac{f(\mathbf{x}) \sum_{i=1}^{n} \alpha_i p_i(\mathbf{x})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})} d\mathbf{x} \tag{22}$$

$$= \int f(\mathbf{x}) d\mathbf{x} \equiv \mu. $$

□

The variance of G is given by

$$V[G] = V\left[\sum_{i=1}^{n} \alpha_i G_i\right] = \sum_{i=1}^{n} \alpha_i^2 V[G_i] = \sum_{i=1}^{n} \frac{\alpha_i^2 \sigma_i'^2}{n_i}. \tag{23}$$

For the sake of the theoretical analysis, we define G^1, a normalized version of G with $N = 1$ (see Section 2.1), with variance

$$V[G^1] = \sum_{i=1}^{n} \frac{\alpha_i^2 \sigma_i'^2}{\beta_i}. \tag{24}$$

Next we study four special cases of estimator G^1.

Remark 1. *We could also consider the one-sample estimator $\mathcal{G}$, randomized version of G. However, $\mathcal{G}$ is a particular case of the general estimator $\mathcal{Z}$, and we have seen in Section 2.2 that the optimal*

case for $\mathcal{Z}$ is when $\mathcal{Z} \equiv \mathcal{F}$, thus it only makes sense to consider the extension G of the multi-sample estimator F.

Remark 2. *Grittmann et al. estimator [19], can be interpreted as a G estimator where $\alpha_i \propto \beta_i \frac{m_i^2}{v_i} = \beta_i \frac{v_i + \mu^2}{v_i}$, where v_i is the variance of i independent technique and $m_i^2 = v_i + \mu^2$ the second moment,*

$$v_i = \int \frac{f(\mathbf{x})^2}{p_i(x)} d\mathbf{x}. \tag{25}$$

3.1. Case 1: $\alpha_i = \beta_i, \forall i$

In this particular case, the estimator G reverts to F. The variance is

$$V[F^1] = \sum_{i=1}^{n} \alpha_i \sigma'^2_i, \tag{26}$$

by simple substitution in Equation (24). We aim at finding the optimal $\{\alpha_i^*\}_{i=1}^{n}$ such the variance of Equation (26) is minimized.

Theorem 2. *The optimal estimator F^* in terms of variance is achieved when $\forall j \in \{1, \dots, n\}$, the following values are equal*

$$\sigma'^2_j + 2\mu'^2_j - 2 \sum_{i=1}^{n} \alpha_i^* \mu_i' \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k^* p_k(\mathbf{x}))^2} d\mathbf{x}. \tag{27}$$

See Appendix B for a proof.

Compare Equation (27) with the condition for optimal $\mathcal{F}^*$, which is [9] that the following expression is equal $\forall j \in \{1, \dots, n\}$,

$$\sigma'^2_j + \mu'^2_j. \tag{28}$$

The optimal $\{\alpha_i^*\}_{i=1}^{n}$ solutions are in general different for F, Equation (27), and for $\mathcal{F}^*$, Equation (28), although observe that, when for $\{\alpha_i^*\}_{i=1}^{n}$ all i the μ_i' are equal, then the optimality condition Equation (27) implies that $\forall i \in \{1, \dots, n\}$ the σ'^2_i are equal too, and thus the optimal solutions for F and $\mathcal{F}$ are the same, in concordance with being $V[F] = V[\mathcal{F}]$ for this special case, see Section 2.2 and Appendix A.

Another interesting result regarding the μ_i' values is the following theorem,

Theorem 3. *If $V[F(\mathcal{F})] = 0$ then for all i, $\mu_i' = \mu$.*

Proof. As $V[F] \le V[F(\mathcal{F})]$ then $V[F] = \sum_i^n \alpha_i \sigma'^2_i = 0$, and thus for all i, $\sigma_i' = 0$; however, $V[F(\mathcal{F})] = 0$ is obviously a local (and global) minimum for $V[F(\mathcal{F})]$ (a convex function) and thus from Equation (28) for all i, $\sigma'^2_i + \mu'^2_i$ are equal and hence the result. Alternatively, we know that $V[F] = V[F(\mathcal{F})] \iff$ for all i, all $\mu_i' = \mu$. □

3.2. Case 2: Fixed $\{\alpha_i\}_{i=1}^{n}$

Consider now that $\{\alpha_i\}_{i=1}^{n}$ are fixed, and hence also $\{\sigma'^2_i\}_{i=1}^{n}$, are fixed. Applying Cauchy–Schwartz inequality to sequences $\{\sqrt{\beta_i}\}_{i=1}^{n}$ and $\{\frac{\alpha_i \sigma'_i}{\sqrt{\beta_i}}\}_{i=1}^{n}$ we obtain

$$\left(\sum_{i=1}^{n} \alpha_i \sigma'_i \right)^2 \le \left(\sum_{i=1}^{n} \beta_i \right) \left(\sum_{i=1}^{n} \frac{\alpha_i^2 \sigma'^2_i}{\beta_i} \right). \tag{29}$$

As the left term is fixed, equality in Equation (29) will give the minimum of right term, being this term $V[G^1]$ as $\sum_{i=1}^{n} \beta_i = 1$; however, equality can only happen when the right

term sequences are proportional, this is, for all i, $\sqrt{\beta_i} \propto \frac{\alpha_i \sigma'_i}{\sqrt{\beta_i}}$, and thus the optimal $\{\beta_i\}_{i=1}^n$ are given by

$$\beta_i^* \propto \alpha_i \sigma'_i, \quad i = 1, \dots n, \tag{30}$$

and the optimal (minimum) variance is

$$V[G^{1*}] = \left(\sum_{i=1}^n \alpha_i \sigma'_i \right)^2. \tag{31}$$

Theorem 4. *Given an estimator F with $\{\alpha_i\}_{i=1}^n$ values, we can always find a better estimator G by sampling as $\beta_i^* \propto \alpha_i \sigma'_i$, which is strictly better whenever not all σ'_i are equal.*

Proof. Observe that, by applying Cauchy–Schwartz inequality to the sequences $\{\sqrt{\alpha_i}\}_{i=1}^n$ and $\{\sqrt{\alpha_i}\sigma'_i\}_{i=1}^n$,

$$\left(\sum_{i=1}^n \alpha_i \sigma'_i \right)^2 \leq \left(\sum_{i=1}^n \alpha_i \right) \left(\sum_{i=1}^n \alpha_i \sigma'^2_i \right), \tag{32}$$

but the left-hand side of inequality is $V[G^{1*}]$, and as $\sum_{i=1}^n \alpha_i = 1$, the right-hand side is $V[F^1]$. Hence, for the optimal values $\{\beta_i^*\}_{i=1}^n$ as in Equation (30), the estimator G^* always outperforms the estimator F (When comparing estimators F and G, we consider, unless explicitly stated, the same set of $\{\alpha_i\}_{i=1}^n$ values).

Equality in Equation (32) only happens when for all i, $\sqrt{\alpha_i} \propto \sqrt{\alpha_i}\sigma'_i$, i.e., when all σ'_i are equal. In that case we have for all i $\beta_i^* = \alpha_i$, and we revert to the F estimator. □

Remark 3. *From the inequality $\left(\sum_{i=1}^n \frac{1}{n}\sigma'^2_i \right) \leq n\left(\sum_{i=1}^n \frac{1}{n}\sigma'_i \right)^2$, the maximum possible acceleration by using the optimal β_i^* values when for all i, $\alpha_i = 1/n$, is equal to n (as observed in [7]). This acceleration would be approached when there is an index k such that for all $i \neq k$, $\sigma'_k >> \sigma'_i$. In general, the more different the σ'_i are, the higher the acceleration.*

A particular case of Theorem 4 is when $\alpha_i = \frac{1}{n}$, $\forall i$, where $V[G^1]$ becomes

$$V[G^1] = \frac{1}{n^2} \sum_{i=1}^n \frac{\sigma'^2_i}{\beta_i}. \tag{33}$$

This case was introduced in Section 4 of [7]. It was shown that this estimator is provably better than F with $\alpha_i = 1/n$, $\forall i$ when

$$\beta_i^* \propto \sigma'_i, \quad i = 1, \dots, n, \tag{34}$$

which is the optimal case of Equation (33). Examples showing the improvement obtained were also given in [7].

Optimal Efficiency

Let us now take into account that the cost of each sampling technique can be different, as it is usually considered in the literature [25]. Let us denote the cost of sampling technique i as c_i. The inverse of efficiency for the estimator G is given by

$$E_G^{-1} = \left(\sum_{i=1}^n \beta_i c_i \right) \left(\sum_{i=1}^n \frac{\alpha_i^2 \sigma'^2_i}{\beta_i} \right). \tag{35}$$

Note that this quantity represents the total cost multiplied by the variance of the estimator. Using Cauchy–Schwartz inequality with the sequences $\{\sqrt{\beta_i c_i}\}_{i=1}^n$ and $\{\frac{\alpha_i \sigma'_i}{\sqrt{\beta_i}}\}_{i=1}^n$, we obtain

$$\left(\sum_{i=1}^n \alpha_i \sigma'_i \sqrt{c_i} \right)^2 \leq \left(\sum_{i=1}^n \beta_i c_i \right) \left(\sum_{i=1}^n \frac{\alpha_i^2 \sigma'^2_i}{\beta_i} \right). \tag{36}$$

The optimal sampling rates (for maximizing the efficiency) are those that yield Equation (36) as an equality, which happens when $\beta_i^* \propto \frac{\alpha_i \sigma'_i}{\sqrt{c_i}}$. Observe that, using again the Cauchy–Schwartz theorem, with sequences $\{\sqrt{\alpha_i c_i}\}_{i=1}^n$ and $\{\sqrt{\alpha_i} \sigma'_i\}_{i=1}^n$, we obtain

$$\left(\sum_{i=1}^n \alpha_i \sigma'_i \sqrt{c_i} \right)^2 \leq \left(\sum_{i=1}^n \alpha_i c_i \right) \left(\sum_{i=1}^n \alpha_i \sigma'^2_i \right), \tag{37}$$

where the left-hand side is E_G^{-1} with the optimal sampling rates, and the right-hand side is E_F^{-1}. Note that equality only happens when for all i, $\sqrt{\alpha_i c_i} = \sqrt{\alpha_i} \sigma'_i$, i.e., $c_i \propto \sigma'^2_i$, where we have $\beta_i^* = \alpha_i$ and we revert to the F estimator. This is summarized in the following theorem.

Theorem 5. *Given an estimator F with $\{\alpha_i\}_{i=1}^n$ values, and sampling costs $\{c_i\}_{i=1}^n$, we can always find a strictly more efficient estimator G^* when for all i, $\beta_i^* \propto \alpha_i \frac{\sigma'_i}{\sqrt{c_i}}$, and not all $c_i \propto \sigma'^2_i$.*

3.3. Case 3: Fixed $\{\beta_i\}_{i=1}^n$

Theorem 6. *Consider now a fixed set $\{\beta_i\}_{i=1}^n$. The optimal set $\{\alpha_i^*\}_{i=1}^n$ can be found using Lagrange multipliers with target function*

$$\Lambda(\{\alpha_i\}_{i=1}^n, \lambda) = \sum_{i=1}^n \frac{\alpha_i^2 \sigma'^2_i}{\beta_i} + \lambda \left(\sum_{i=1}^n \alpha_i - 1 \right).$$

Observe that the σ'^2_i values depend on the $\{\alpha_i\}_{i=1}^n$ values. The optimal values are those that obey, for all j, the following expression

$$\frac{\alpha_j^* \sigma'^2_j}{\beta_j} = \sum_{i=1}^n \frac{\alpha_i^{*2}}{\beta_i} \times$$
$$\left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k^* p_k(\mathbf{x}))^3} d\mathbf{x} - \mu'_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k^* p_k(\mathbf{x}))^2} d\mathbf{x} \right). \tag{38}$$

Proof. The derivation can be found in the Appendix C. □

Note that in the general case, the optimal $\alpha_i^* \neq \beta_i$.

Moreover, aside from the optimal values $\{\alpha_j^*\}_{j=1}^n$, we can find cases where $V[G^1] \leq V[F^1]$. Given $\{\beta_i\}_{i=1}^n$, from Theorem 4 we have that if there exist $\{\alpha_i\}_{i=1}^n$ such that $\beta_i \propto \alpha_i \sigma'_i$, $i = 1, ...n$, then $V[G^1] \leq V[F^1]$. From Theorem 5, we have that if there exist $\{\alpha_i\}_{i=1}^n$ such that $\beta_i \propto \alpha_i \frac{\sigma'_i}{\sqrt{c_i}}$, $\forall i = 1, ...n$, then the estimator G^1 is more efficient than estimator F^1.

Observe that from Equation (38), a necessary and sufficient condition for the optimal solution α_i^* to be such that $\alpha_i^* = \beta_i$ for all i is that Equation (39) holds (see Appendix C),

$$\mu'^2_j = \sum_{i=1}^n \alpha_i \mu'_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x}, \tag{39}$$

which is satisfied when for all i, $\mu'_i = \mu$. In this particular case, $\mu'_i = \mu$, $G \equiv F$, and $V[G] = V[F] = V[\mathcal{F}]$. In other words, when choosing a sampling rate of $\{\beta_i\} = \{\alpha_i^\star\}$,

where $\{\alpha_i^\star\}$ are such that for all i, $\mu_i' = \mu$, the optimal $\{\alpha_i\}$ parameters for G estimator are precisely $\{\alpha_i^\star\}$. That is, we can not improve the $F(\{\alpha_i^\star\})$ estimator with the G estimator with a suitable selection of $\{\beta_i\}$ parameters. Observe that it is similar to the case studied in Section 3.2 for the $\{\alpha_i\}$ that make all σ'_i equal.

Equation (39) might have additional solutions to the one (if exists) that makes all μ_i' equal, in this case $V[G] = V[F]$ too but $V[F] \neq V[\mathcal{F}]$.

3.4. Case 4: $\beta_i = 1/n, \forall i$

In the case when for all i, $\beta_i = 1/n$, the variance becomes

$$V[G^1] = \sum_{i=1}^n \frac{\alpha_i^2 \sigma'^2_i}{1/n} = n \sum_{i=1}^n \alpha_i^2 \sigma'^2_i. \tag{40}$$

Note that this is a usual case in the MIS literature strategies [2,4–6] and the adaptive IS (AIS) literature [11–15,24], since all the techniques have the same number of counts. By setting in Equation (38) for all i, $\beta_i = 1/n$, and if we can optimize $\{\alpha_j\}_{j=1}^n$, we can find the minimum variance values $\{\alpha_j^*\}_{j=1}^n$. Thus the minimum variance $V[G^*]$ corresponds, if the solution exists, to the values $\{\alpha_j^*\}_{j=1}^n$ that satisfy

$$\alpha_j^* \sigma'^2_j = \sum_{i=1}^n \alpha_i^{*2} \tag{41}$$
$$\times \left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k^* p_k(\mathbf{x}))^3} d\mathbf{x} - \mu_i' \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k^* p_k(\mathbf{x}))^2} d\mathbf{x} \right).$$

The corresponding variance $V[G^*]$ will be less or equal than the variance of $V[G]$ for all $\{\alpha_j\}_{j=1}^n$, and in particular for $\alpha_i = 1/n$, where G converts into F as $\beta_i = \alpha_i = 1/n$, the classic balance heuristic estimator, thus $V[G^*(\{\alpha_i^*\}, \{\beta_i = 1/n\})] \leq V[F(\{\alpha_i = 1/n\})]$. Observe that in general $\{\alpha_i^* \neq 1/n\}$, because substituting $\{\alpha_i^* = 1/n\}$ in Equation (41) the resulting equation Equation (39) is not satisfied in general.

Comparing $V[G^1(\{\alpha_i\}, \{\beta_i\})]$, $V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\})]$ and $V[F^1(\{\alpha_i\})]$

Let us see first when $V[G^1(\{\alpha_i\}, \{\beta_i\})] \leq V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\}))]$ and viceversa.

Theorem 7. *If the sequences $\{\beta_i\}$, $\{\frac{\alpha_i^2 \sigma'^2_i}{\beta_i}\}$ are comonotonic, i.e., $\beta_i \leq \beta_j \implies \frac{\alpha_i^2 \sigma'^2_i}{\beta_i} \leq \frac{\alpha_j^2 \sigma'^2_j}{\beta_j}$, then*

$$V[G^1(\{\alpha_i\}, \{\beta_i\})] \leq V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\})]. \tag{42}$$

Proof. Applying Theorem 9, Corollary 5 from [26],

$$\sum_{i=1}^n \frac{1}{n} \left(\frac{\alpha_i^2 \sigma'^2_i}{\beta_i} \right) \quad \leq \quad \sum_{i=1}^n \beta_i \left(\frac{\alpha_i^2 \sigma'^2_i}{\beta_i} \right), \tag{43}$$

obtaining the inequality in Equation (42). See also [10,27]. □

Observe that equality in Equation (43) happens when for all i, $\beta_i \propto \alpha_i^2 \sigma'^2_i$.

In the same way we can proof the inverse case as follows,

Theorem 8. *If the sequences $\{\beta_i\}$ and $\{\frac{\alpha_i^2\sigma_i'^2}{\beta_i}\}$ are countermonotonic, i.e., $\beta_i \leq \beta_j \implies \frac{\alpha_i^2\sigma_i'^2}{\beta_i} \geq \frac{\alpha_j^2\sigma_j'^2}{\beta_j}$, then*

$$V[G^1(\{\alpha_i\}, \{\beta_i\})] \geq V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\}))]. \tag{44}$$

The proof implies for countermonotonicity a change of sign in the inequality in Theorem 9, Corollary 5 from [26].

Let us compare now $V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\})]$ and $V[F^1(\{\alpha_i\})]$.

Theorem 9. *If the sequences $\{\alpha_i\}$, $\{\alpha_i\sigma_i'^2\}$ are countermonotonic, i.e., $\alpha_i \leq \alpha_j \implies \alpha_i\sigma_i'^2 \geq \alpha_j\sigma_j'^2$, then*

$$V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\})] \leq V[F^1(\{\alpha_i\})]. \tag{45}$$

Proof. The proof is as above, making use of Theorem 9, Corollary 5 from [26],

$$\sum_{i=1}^{n}\frac{1}{n}(\alpha_i\sigma_i'^2) \geq \sum_{i=1}^{n}\alpha_i(\alpha_i\sigma_i'^2), \tag{46}$$

obtaining the inequality in Equation (45). □

Observe that equality in Equation (46) happens when for all i, $\alpha_i \propto 1/\sigma_i'^2$. In the same way we can prove the reverse case of Theorem 9.

Theorem 10. *If the sequences $\{\alpha_i\}$, $\{\alpha_i\sigma_i'^2\}$ are comonotonic, i.e., $\alpha_i \leq \alpha_j \implies \alpha_i\sigma_i'^2 \leq \alpha_j\sigma_j'^2$, then*

$$V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\})] \geq V[F^1(\{\alpha_i\})]. \tag{47}$$

As a direct consequence of Theorems 9 and 10 we have the following corollary:

Corollary 1. *When $\alpha_i \propto 1/\sigma_i'^{2+\epsilon}$, if $-2 \leq \epsilon \leq 0$ then $V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\})] \leq V[F^1(\{\alpha_i\})]$ holds, and if $\epsilon \geq 0$ then $V[G^1(\{\alpha_i\}, \{\beta_i = 1/n\})] \geq V[F^1(\{\alpha_i\})]$ holds.*

By setting $\epsilon = -1$ in Corollary 2, the products $\alpha_i\sigma_i'$ are equal for all i, which is equivalent to saying that $\alpha_i\sigma_i' \propto 1/n = \beta_i$. Observe that this corresponds to the optimal solution $\{\beta_j^*\}$ given by Equation (30). Observe now that the inequality for this optimal solution, $V[G^1] \leq V[F^1]$, can be obtained by successive application of Theorems 7 and 9.

Theorem 11. *If the sequences $\{\alpha_i\}$, $\{\alpha_i\sigma_i'^2\}$ are countermonotonic, and the sequences $\{\beta_i\}$, $\{\frac{\alpha_i^2\sigma_i'^2}{\beta_i}\}$ are comonotonic, then*

$$V[G^1(\{\alpha_i\}, \{\beta_i\})] \leq V[F^1(\{\alpha_i\})]. \tag{48}$$

Proof. The result is obtained by successive application of Theorems 7 and 9. □

Theorem 12. *If the sequences $\{\alpha_i\}$, $\{\alpha_i\sigma_i'^2\}$ are comonotonic, and the sequences $\{\beta_i\}$, $\{\frac{\alpha_i^2\sigma_i'^2}{\beta_i}\}$ are countermonotonic, then*

$$V[G^1(\{\alpha_i\}, \{\beta_i\})] \geq V[F^1(\{\alpha_i\})]. \tag{49}$$

As a direct consequence of Theorems 11 and 12 we have the following corollary:

Corollary 2. *When $\alpha_i \propto 1/\sigma'^{2+\epsilon}_i$ and $\beta_i \propto (\alpha_i\sigma'_i)^\gamma$, if $-2 \leq \epsilon \leq 0$ and $0 < \gamma \leq 2$ then $V[G^1(\{\alpha_i\},\{\beta_i\})] \leq V[F^1(\{\alpha_i\})]$ holds, and if $\epsilon \geq 0$ and $\gamma > 2$ then $V[G^1(\{\alpha_i\},\{\beta_i\})] \geq V[F^1(\{\alpha_i\})]$ holds.*

For example, take $\alpha_i \propto 1/\sigma'^2_i$, and $\beta_i \propto \alpha_i\sigma'_i \propto 1/\sigma'_i$, then $V[G^1(\{\alpha_i\},\{\beta_i\})] \leq V[F^1(\{\alpha_i\})]$.

Finally, observe that we do not have a sufficient condition for $V[F^1(\{\alpha_i\})] \leq V[F^1(\{\alpha_i = 1/n\})]$, only there are heuristics [7,9,10,28].

3.5. Case 5: General Case

Let us consider the global optimum for G estimator, this is, without imposing any constraint to the parameters $\{\alpha_i\}, \{\beta_i\}$. For this we should optimize the lagrangian function

$$\Lambda(\{\alpha_i\}^n_{i=1},\{\beta_i\}^n_{i=1},\lambda_1,\lambda_2) = \sum_{i=1}^n \frac{\alpha_i^2\sigma'^2_i}{\beta_i} + \lambda_1\left(\sum_{i=1}^n \alpha_i - 1\right) + \lambda_2\left(\sum_{i=1}^n \beta_i - 1\right).$$

Differentiating with respect to α_i,

$$\frac{\partial\Lambda(\{\alpha_i\}^n_{i=1},\{\beta_i\}^n_{i=1},\lambda_1,\lambda_2)}{\partial\alpha_j} = 0,$$

we obtain Equation (38).

Differentiating with respect to β_i,

$$\frac{\partial\Lambda(\{\alpha_i\}^n_{i=1},\{\beta_i\}^n_{i=1},\lambda_1,\lambda_2)}{\partial\beta_j} = 0,$$

we obtain Equation (30). Thus, the optimal parameters $\{\alpha_i^\star\}^n_{i=1}$ and $\{\beta_i^\star\}^n_{i=1}$ will solve for Equations (38) and (30) simultaneously.

Let us consider now the $\{\alpha_i^\dagger\}^n_{i=1}$, if they exist, such that for all i, σ'_i are equal. Then, for Equation (30) to hold for $\{\alpha_i^\dagger\}^n_{i=1}$, we have that $\beta_i^\dagger = \alpha_i^\dagger$ for all $i = 1,\ldots,n$. From Equation (38), taking $\{\beta_i^\dagger\}^n_{i=1} = \{\alpha_i^\dagger\}^n_{i=1}$, Equation (39) has to hold for all $\alpha_i^\dagger$.

Furthermore, the holding of Equation (39), together with the hypothesis that for all j, σ'^2_j are equal, allows Equation (27) to hold, and then $\{\alpha_i^\dagger\}^n_{i=1}$ is an optimum for F too, $V[F] = V[G]$. If in addition, if $\{\alpha_i^\dagger\}$ are such that they make all $\{\mu'\}^n_{j=1}$ equal, then they would be optimal for $\mathcal{F}$ too, $V[\mathcal{F}] = V[F] = V[G]$.

4. Singular Solutions

The variance of $\mathcal{F}$ can be written as [7]

$$\begin{aligned}V[\mathcal{F}] &= \sum_{i=1}^n \alpha_i(\sigma'^2_i + \mu'^2_i - \mu^2) = \sum_{i=1}^n \alpha_i\sigma'^2_i + \sum_{i=1}^n \alpha_i\mu'^2_i - \mu^2. \qquad (50)\\ &= \sum_{i=1}^n \alpha_i(\sigma'^2_i + \mu'^2_i) - \mu^2.\end{aligned}$$

The optimal parameters $\{\alpha_i\}^n_{i=1}$ happen when for all i, $\sigma'^2_i + \mu'^2_i$ equal, which gives a local minimum for $V[\mathcal{F}]$ (in this case a global minimum too, as $V[\mathcal{F}]$ is convex). Observe that this makes, in the third inequality, all terms factored by $\{\alpha_i\}$ equal, and also corresponds to equality in the Cauchy–Schwartz inequality

$$\left(\sum_{i=1}^n \alpha_i\sqrt{(\sigma'^2_i + \mu'^2_i)}\right)^2 \leq \left(\sum_{i=1}^n \alpha_i\right)\left(\sum_{i=1}^n \alpha_i(\sigma'^2_i + \mu'^2_i)\right), \qquad (51)$$

with equality only when, for all i, $\sigma'^2_i + \mu'^2_i$ are equal.

If there exist $\{\alpha_i^\star\}_{i=1}^n$ values such that for all i, $\mu_i' = \mu$ and all σ'_i are equal, these values correspond to the global optimum $V[\mathcal{F}^\star] = V[F^\star] = V[G^\star]$, and $F^\star \equiv G^\star$, but $F^\star \not\equiv \mathcal{F}^\star$.

Suppose there exist $\{\alpha_i^\star\}_{i=1}^n$ such that for all i, $\mu_i' = \mu$, then we can not improve the F estimator with a G estimator with a suitable selection of $\{\beta_i\}_{i=1}^n$ parameters, because the optimal $\{\alpha_i\}$ parameters for the G estimator when $\{\beta_i\}_{i=1}^n = \{\alpha_i^\star\}_{i=1}^n$ are precisely $\{\alpha_i^\star\}$. This is, the optimal of $V[G(\{\alpha_i\}_{i=1}^n, \{\beta_i\}_{i=1}^n = \{\alpha_i^\star\}_{i=1}^n)]$ is for $\{\alpha_i\}_{i=1}^n = \{\alpha_i^\star\}_{i=1}^n$.

Let us see now $V[F]$:

$$V[F] = \sum_{i=1}^n \alpha_i \sigma'^2_i.$$

Suppose there exist $\{\alpha_i^\star\}_{i=1}^n$ such that all σ'_i are equal, then we can not improve the F estimator with a suitable selection of $\{\beta_i\}_{i=1}^n$ parameters for the G estimator, because the optimal $\{\beta_i^\star\}_{i=1}^n$ are such that $\{\beta_i^\star\}_{i=1}^n = \{\alpha_i^\star\}_{i=1}^n$. This is, the optimal of $V[G(\{\alpha_i^\star\}_{i=1}^n, \{\beta_i\}_{i=1}^n)]$ is for $\{\beta_i\}_{i=1}^n = \{\alpha_i^\star\}_{i=1}^n$. Observe that by Cauchy–Schwartz

$$\left(\sum_{i=1}^n \alpha_i \sigma'_i \right)^2 \le \left(\sum_{i=1}^n \alpha_i \right) \left(\sum_{i=1}^n \alpha_i \sigma'^2_i \right), \tag{52}$$

with equality only when, for all i, all the σ'^2_i are equal. Observe thus that both the solutions for μ_i' equal and for σ'^2_i equal are singular solutions for G, where it reverts to F.

Let us consider now $V[G]$:

$$V[G] = \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \sigma'^2_i = \sum_{i=1}^n \alpha_i \left(\frac{\alpha_i}{\beta_i} \sigma'^2_i \right) = \sum_{i=1}^n \beta_i \left(\frac{\alpha_i^2}{\beta_i^2} \sigma'^2_i \right). \tag{53}$$

In the second equality in Equation (53), the solution for all i, $\frac{\alpha_i}{\beta_i} \sigma'^2_i$ equal, or $\beta_i \propto \alpha_i \sigma'^2_i$, makes $V[G] = V[F]$, because then for all i, $\alpha_i \sigma'^2_i = K \beta_i$, and substituting in the expressions for $V[G], V[F]$ we have $V[G] = V[F] = K$, but $G \not\equiv F$. By Cauchy–Schwartz,

$$\left(\sum_{i=1}^n \sqrt{\frac{\alpha_i^3 \sigma'^2_i}{\beta_i}} \right)^2 \le \left(\sum_{i=1}^n \alpha_i \right) \left(\sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \sigma'^2_i \right), \tag{54}$$

with equality only when, for all i, $\frac{\alpha_i}{\beta_i} \sigma'^2_i$ are equal.

In the third equality in Equation (53), the solution for all i, $\frac{\alpha_i^2}{\beta_i^2} \sigma'^2_i$, equal, or $\beta_i \propto \alpha_i \sigma'_i$, gives the optimal solution for $V[G]$ when $\{\alpha_i\}$ are fixed, Equation (30). By Cauchy–Schwartz,

$$\left(\sum_{i=1}^n \alpha_i \sigma'_i \right)^2 \le \left(\sum_{i=1}^n \beta_i \right) \left(\sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \sigma'^2_i \right), \tag{55}$$

with equality only when, for all i, $\frac{\alpha_i^2}{\beta_i^2} \sigma'^2_i$ are equal.

We summarize in Table 2 the different possibilities.

Table 2. Singular parameter values. $\{\alpha_i^\star\}_{i=1}^n$, $\{\beta_i^\star\}_{i=1}^n$ are solutions of the equations in first column and same row.

for all i, $\sigma_i'^2 + \mu_i'^2$ equal	$V[\mathcal{F}(\{\alpha_i^\star\})]$ global minimum
for all i, $\sigma_i'^2$ equal and μ'_i equal	$V[\mathcal{F}(\{\alpha_i^\star\})] = V[F(\{\alpha_i^\star\})] = V[G(\{\alpha_i^\star\}, \{\beta_i\} = \{\alpha_i^\star\})]$ global minimum, $F \equiv G$, but $F \not\equiv \mathcal{F}$
for all i, μ'_i equal	$V[\mathcal{F}(\{\alpha_i^\star\})] = V[F(\{\alpha_i^\star\})]$; optimal of $V[G(\{\alpha_i\}, \{\beta_i\} = \{\alpha_i^\star\})]$ is for $\{\alpha_i\} = \{\alpha_i^\star\}$
for all i, σ'_i equal	optimal of $V[G(\{\alpha_i^\star\}, \{\beta_i\})]$ is for $\{\beta_i\} = \{\alpha_i^\star\}$, $F \equiv G$
for all i, $\frac{\alpha_i}{\beta_i}\sigma'^2_i$ equal, or $\beta_i \propto \alpha_i\sigma'^2_i$	$V[G(\{\alpha_i^\star\}, \{\beta_i^\star\})] = V[F(\{\alpha_i^\star\})]$
for all i, $\frac{\alpha_i^2}{\beta_i^2}\sigma'^2_i$, equal, or $\beta_i \propto \alpha_i\sigma'_i$	optimal of $V[G(\{\alpha_i\}, \{\beta_i\})]$ when $\{\alpha_i\}$ are fixed, $\forall\{\alpha_i\}$, $V[G(\{\alpha_i\}, \{\beta_i^\star\})] \leq V[F(\{\alpha_i\})]$

5. Relationship with χ^2 Divergence: A Necessary and Sufficient Condition for $V[G] \leq V[F]$

The variance of an importance sampling estimator can be written in terms of a χ^2 divergence if $f(x) \geq 0$ [29,30]. As the $\mathcal{F}$ estimator can be seen as an importance sampling estimator with pdf $p(\mathbf{x}) = \sum_i^n p_i(\mathbf{x})$, its variance can be written as

$$
\begin{aligned}
V[\mathcal{F}] &= \int \frac{f(\mathbf{x})^2}{\sum_i^n \alpha_i p_i(\mathbf{x})} d\mathbf{x} - \mu^2 \\
&= \mu^2 \int \frac{\frac{f(\mathbf{x})^2}{\mu^2}}{\sum_i^n \alpha_i p_i(\mathbf{x})} d\mathbf{x} - 2\mu^2 \int \frac{f(\mathbf{x})}{\mu} d\mathbf{x} + \mu^2 \int \left(\sum_i^n \alpha_i p_i(\mathbf{x})\right) d\mathbf{x} \\
&= \mu^2 \int \frac{\left(\frac{f(\mathbf{x})}{\mu} - (\sum_i^n \alpha_i p_i(\mathbf{x}))\right)^2}{\sum_i^n \alpha_i p_i(\mathbf{x})} d\mathbf{x} \\
&= \mu^2 \chi^2\left(\frac{f(\mathbf{x})}{\mu}, \sum_i^n \alpha_i p_i(\mathbf{x})\right).
\end{aligned}
\tag{56}
$$

Observe that from the χ^2 expression in Equation (56) it is clear that $V[\mathcal{F}] = 0 \iff f(\mathbf{x}) \propto \sum_i^n \alpha_i p_i(\mathbf{x})$ except possibly in a zero measure set. As a consequence, for all i $\mu_i' = \mu$, and thus $V[F] = V[\mathcal{F}] = 0$. As $V[F] \leq V[\mathcal{F}]$, we have the result

$$
V[F] = 0 \iff V[\mathcal{F}] = 0.
\tag{57}
$$

A possible generalization of Equation (56) to the $\mathcal{G}$ estimator, the randomized version of G, is given in Appendix D.

Although neither $V[F]$ nor $V[G]$ can be written as in Equation (56), we can still relate them to χ^2 divergence. In general, $\sum_i^n \alpha_k \sigma'_k > 0$. Note that, if $\sum_k^n \alpha_k \sigma'_k = 0$, then, for all k, $\sigma'_k = 0$ and $V[F] = V[G] = 0$. Thus both $V[G], V[F] = 0$ or both $V[G], V[F] > 0$. Let us denote $K = \sum_{k=1}^n \alpha_k \sigma'_k$. Then,

$$
\begin{aligned}
V[G] &= \sum_{i=1}^{n} \frac{\alpha_i^2 \sigma'^2_i}{\beta_i} \\
&= K^2 \sum_{i=1}^{n} \frac{\frac{\alpha_i^2 \sigma'^2_i}{K^2}}{\beta_i} + K^2 \sum_{i=1}^{n} \beta_i - K^2 + K^2 \sum_{i=1}^{n} 2\frac{\alpha_i \sigma'_i}{K} - K^2 \sum_{i=1}^{n} 2\frac{\alpha_i \sigma'_i}{K} \\
&= K^2 \left(\sum_{i=1}^{n} \frac{\frac{\alpha_i^2 \sigma'^2_i}{K^2} - 2\frac{\alpha_i \sigma'_i}{K}\beta_i + \beta_i^2}{\beta_i} \right) - K^2 + K^2 \sum_{i=1}^{n} 2\frac{\alpha_i \sigma'_i}{K} \\
&= K^2 \left(\sum_{i=1}^{n} \frac{\frac{\alpha_i^2 \sigma'^2_i}{K^2} - 2\frac{\alpha_i \sigma'_i}{K}\beta_i + \beta_i^2}{\beta_i} \right) + K^2 \\
&= K^2 \left(\sum_{i=1}^{n} \frac{(\frac{\alpha_i \sigma'_i}{K} - \beta_i)^2}{\beta_i} + 1 \right) \\
&= \left(\sum_{k=1}^{n} \alpha_k \sigma'_k \right)^2 \left(\chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\beta_i\} \right) + 1 \right).
\end{aligned}
\tag{58}
$$

For $\{\beta_i \propto \alpha_i \sigma'\}_{i=1}^{n}$ the χ^2 divergence value is zero, and $V[G] = (\sum_{k=1}^{n} \alpha_k \sigma'_k)^2$. This is the optimal of $V[G]$ when the $\{\alpha_i\}_{i=1}^{n}$ are fixed, as seen in Section 3.2. Substituting $\{\beta_i = \alpha_i\}_{i=1}^{n}$ in Equation (58) we obtain

$$
V[F] = \left(\sum_{k=1}^{n} \alpha_k \sigma'_k \right)^2 \left(\chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\alpha_i\} \right) + 1 \right).
\tag{59}
$$

As we have taken $\sum_{k=1}^{n} \alpha_k \sigma'_k > 0$, we can divide first and last terms of Equations (58) and (59),

$$
V[G] = V[F] \frac{\chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\beta_i\} \right) + 1}{\chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\alpha_i\} \right) + 1}.
\tag{60}
$$

Hence the following theorem,

Theorem 13. *If $V[G], V[F] > 0$, then*

$$
V[G] \leq V[F] \iff \chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\beta_i\} \right) \leq \chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\alpha_i\} \right).
$$

Observe that Theorem 13 does not impose any a priori condition on the $\{\alpha_i\}, \{\beta_i\}$ parameters and generalizes Theorem 12. In the same way we can generalize Theorems 7–11 with the following corollaries,

Corollary 3. *If $V[G], V[G(\{\beta_i = 1/n\})] > 0$, then*

$$
V[G] \leq V[G(\{\beta_i = 1/n\})] \iff \chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\beta_i\} \right) \leq \chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{1/n\} \right).
$$

Corollary 4. *If $V[G(\{\beta_i = 1/n\})], V[F] > 0$, then*

$$
V[G(\{\beta_i = 1/n\})] \leq V[F] \iff \chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{1/n\} \right) \leq \chi^2\left(\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^{n} \alpha_k \sigma'_k}\}, \{\alpha_i\} \right).
$$

Remark 4. *Observe that, using the second order χ^2 approximation of KL (Kullback–Leibler) distance between two distributions X_1, X_2, $\chi^2(X_1, X_2) \approx 2KL(X_1, X_2)$ [31], Theorem 13 can be written as*

$$
\begin{aligned}
V[G] \quad &\leq \quad V[F] \iff KL\left(\left\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^n \alpha_k \sigma'_k}\right\}, \{\beta_i\}\right) \leq KL\left(\left\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^n \alpha_k \sigma'_k}\right\}, \{\alpha_i\}\right) \\
&\iff \quad CE\left(\left\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^n \alpha_k \sigma'_k}\right\}, \{\beta_i\}\right) \leq CE\left(\left\{\frac{\alpha_i \sigma'_i}{\sum_{k=1}^n \alpha_k \sigma'_k}\right\}, \{\alpha_i\}\right),
\end{aligned}
$$

where cross entropy $CE(X_1, X_2) = H(X_1) + KL(X_1, X_2)$, and H stands for entropy.

The same approximation can be used for Corollaries 3 and 4.

6. Numerical Examples

6.1. Efficiency Comparison between F and G Estimators

We compare the efficiencies for the *F* estimator and the optimal *G* estimator defined in Theorem 5 with five examples. Table 3 shows the inverse of the efficiencies, $E_F^{-1} = V[F] \cdot Cost[F]$ and $E_G^{-1} = V[G] \cdot Cost[G]$, i.e., the product of variance and cost for the *F* estimator and for the optimal *G* estimator, for these possible sets of $\{\alpha_k\}_{k=1}^n$: (i) equal count of samples, $\{\alpha_k = 1/n\}$; (ii) count of samples inversely proportional to the variances of the independent techniques times the cost of sampling the technique, $\{\alpha_k \propto \frac{1}{c_k v_k}\}$ [8,9]; and the three balance heuristic estimators defined in [28], which are (iii) count of samples inversely proportional to the second moments of the independent techniques times the cost of sampling the technique, $\{\alpha_k \propto \frac{1}{c_k m_k^2}\}$; (iv) count of samples proportional to $\sigma_{k,eq} = \frac{\sigma'_k(\{\alpha_i=1/n\}_{i=1}^n)}{n}$, and inversely proportional to the square root of the cost, $\{\alpha_k \propto \frac{\sigma_{k,eq}}{\sqrt{c_k}}\}$; and

(v) $\{\alpha_k \propto \frac{M_{k,eq}}{\sqrt{c_k}}\}$ where $M_{k,eq} = \sqrt{\int \frac{f^2(x) p_k(x)}{(\sum_i p_i(x))^2}}$.

Below, we describe the five examples. From the results in Table 3, we can see:

- As expected, Theorems 4 and 5 hold for all the cases.
- Examples 1–4 show a general gain in efficiency, around 20%.
- Example 5 with equal costs shows a gain in efficiency from 2% to around 20%.
- Example 5 with different costs shows big gains in efficiency, in particular for equal count of sampling the efficiency is doubled.

We see in all our examples an increase in efficiency when the estimator *G* is used instead of the estimator *F*, showing in some cases, as in Example 5, a 50% improvement. Theorems 4 and 5 ensure that estimator *G* is always better than estimator *F* if the optimal $\{\beta_i\}$ values are used; however, for equal costs (i.e., comparing only variances), the amount of improvement strongly depends on the example considered, from very small o negligible (see first three rows of Table 4 for the first four examples) to an important one (compare Table 3 for Example 5 with equal costs with the first three rows of Table 4 for Example 5). The values are computed for the results in Tables 3 and 4 by using numerical approximations obtained with Mathematica software. It would also be possible to select them in an adaptive and iterative way. We defer the investigation of these adaptive schemes for a future work.

6.1.1. Example 1

Suppose we want to solve the integral

$$
\mu = \int_{\frac{3}{2\pi}}^{\pi} \mathbf{x}\left(\mathbf{x}^2 - \frac{\mathbf{x}}{\pi}\right) \sin(\mathbf{x}) d\mathbf{x} \approx 10.29 \tag{61}
$$

by MIS sampling on functions $\mathbf{x}$, $(\mathbf{x}^2 - \frac{\mathbf{x}}{\pi})$, and $\sin(\mathbf{x})$, respectively. We first find the normalization constants: $\int_{\frac{3}{2\pi}}^{\pi} \mathbf{x} d\mathbf{x} = 4.82082$, $\int_{\frac{3}{2\pi}}^{\pi} (\mathbf{x}^2 - \frac{\mathbf{x}}{\pi}) d\mathbf{x} = 8.76463$, $\int_{\frac{3}{2\pi}}^{\pi} \sin(\mathbf{x}) d\mathbf{x} = 1.88816$. The costs for sampling the techniques are $(1; 6.24; 3.28)$.

6.1.2. Example 2

Let us solve the integral

$$\mu = \int_{\frac{3}{2\pi}}^{\pi} \left(\mathbf{x}^2 - \frac{\mathbf{x}}{\pi} \right) \sin^2(\mathbf{x}) d\mathbf{x} \approx 3.60 \tag{62}$$

using the same functions $\mathbf{x}$, $(\mathbf{x}^2 - \frac{\mathbf{x}}{\pi})$, and $\sin(\mathbf{x})$ as before.

6.1.3. Example 3

As the third example, let us solve the integral

$$\mu = \int_{\frac{3}{2\pi}}^{\pi} \mathbf{x} + \left(\mathbf{x}^2 - \frac{\mathbf{x}}{\pi} \right) + \sin(\mathbf{x}) d\mathbf{x} \approx 15.47 \tag{63}$$

using the same functions as before. The optimal α values for this example, with zero variance, are $\frac{1}{4.82082 + 8.76463 + 1.88816}(4.82082, 8.76463, 1.88816)$.

6.1.4. Example 4

As the fourth example, consider the integral of the sum of the three pdfs

$$\mu = \int_{\frac{3}{2\pi}}^{\pi} 30 \frac{\mathbf{x}}{4.82082} + 30 \frac{(x^2 - \frac{\mathbf{x}}{\pi})}{8.76463} + 40 \frac{\sin(\mathbf{x})}{1.88816} d\mathbf{x} \tag{64}$$
$$\approx 100.$$

In this case we know again the optimal (zero variance) α values: $(0.3, 0.3, 0.4)$. The difference with the previous example is that this case should be most favorable to equal count of samples.

6.1.5. Example 5

As the last example, consider solving the following integral

$$\mu = \int_{0.01}^{\pi/2} (\sqrt{\mathbf{x}} + \sin \mathbf{x}) d\mathbf{x} \approx 2.31175. \tag{65}$$

by MIS sampling on functions $2 - \mathbf{x}$, and $\sin^2(\mathbf{x})$.

6.2. *Variances of Examples 1–5 for Some Notable Cases*

In Table 4 we present some notable cases for the variance of Examples 1–5. Observe that here we do not consider the cost of sampling. The values are computed for the results in Table 4 with Mathematica software. We have

- In first, second and third row we present the optimal values of $V[\mathcal{F}] \geq V[F] \geq V[G]$, respectively. In the first example those values are equal up to the fourth decimal position. In the second example they are equal up to the second decimal position, in the third and fourth examples they are equal, and in the fifth example the gain of $V[G]$ with respect to $V[F]$ is more than 30%.
- In the fourth row we present the values of $V[F(1/n)]$, which are compared against several strategies and heuristics in the following rows
- The values in the fifth row for $V[G]$, that correspond to Equation (30) (or the left-hand side of Equation (37) with all costs equal to 1), although being better than the values

- for $V[F]$ in the fourth row, as expected, do not give any significant improvement. As we have seen in Table 3 the improvements for this case come rather from efficiency.
- The values in the sixth row for $V[G]$, that correspond to the solution of Equation (40), this is, the optimal $\{\alpha_i\}$ for $\{\beta_i = 1/n\}$, are smaller than the variances of $V[F]$ for equal sampling rate, fourth row, and much smaller in some of the examples, with variances equal (as in Examples 3 and 4), or less, as in Example 5, than the optimal $V[F]$.
- Although in general there is no $\{\alpha_i\}$ solution for all μ_i' equal, for instance the numerical approximation obtained in Example 1 is $\mu_1' = 10.2624$, $\mu_2' = \mu_3' = 10.2876$, the variances obtained, in the seventh row of Table 4, are for all the examples very near to the optimal value of $V[F]$. Observe that if these $\{\alpha_i^\star\} = arg\ \forall i, \mu_i'\ equal$ values exist, then they are the optimal $\{\alpha_i\}$ values for $V[G(\beta_i = \alpha_i^\star)]$, i.e., for the sampling rates $\{\beta_i = \alpha_i^\star\}$ the G estimator can not improve on the F estimator.
- In row 8 we show the G estimator corresponding $\alpha_i\sigma_i' \approx equal$, with $\beta_i = 1/n$. It improves on $F(1/n)$ for all examples except the third one, where it scores closely, beating estimators in rows 9 and 10 except for Example 3. It is indeed a theoretical estimator as there is no easy way to solve this equation. However, it can be approximated by the heuristic in line 11, see below. Observe that if $\{\alpha_i^\star\}$ are the solutions of for all i, $\alpha_i\sigma_i'$ equal, then the optimal $\{\beta_i\}$ for $G(\{\alpha_i^\star\}, \{\beta_i\})$ are $\{\beta_i = 1/n\}$.
- The results in row 9 correspond to the estimator defined in [19], $\alpha_i \propto m_i^2/v_i$, where $m_i^2 = v_i + \mu^2$ are the second moments of the independent techniques, and $\beta_i = 1/n$. The results are slightly better than $F(1/n)$ for Examples 1 and 2, much better for Example 3, much worse for Example 4 and slightly worse for Example 5. Observe that the m_i, v_i values can be easily approximated using a first batch of samples in a Monte Carlo integration, as it was already performed in [7,9,10].
- In row 10 we introduce the estimator corresponding to $\alpha_i \propto 1/v_i$, and $\beta_i = 1/n$, which gives better results than $F(1/n)$ for Examples 1 and 2, much better for Example 3, worse for Example 5, and much worse for Example 4. This estimator improves on the one in [19] except for Example 4. Observe that these α_i weights correspond to the optimal weights in the linear combination of estimators when the sampling rates (β_i) are fixed, Equation (13) of [9].
- In row 11 we approximate the estimator in row 8 by using $\{\sigma_{i,1/n}'\}$, which corresponds to $\{\alpha_i = 1/n\}$, to approximate $\{\sigma_i'\}$. Observe that $\{\sigma_{i,1/n}'\}$ values can be easily obtained with a first batch of samples in a Monte Carlo integration. In all cases, except in Example 3, we improve on $F(1/n)$. The big error on Example 4 of the estimator in lines 9 and 10 is controlled.
- In row 12 we modulate the estimator in row 11 with the μ_i' values. Observe that now for all the examples the variance is less than for $F(1/n)$.
- The results in row 13 show that, even if $F(\{\alpha_i\}) \leq F(\{1/n\})$, we can not guarantee that $G(\{\alpha_i\}, \{\beta_i = 1/n\}) \leq F(\{1/n\})$.

From the results in Table 4 row 6, we can conclude that there can be room for improvement on $F(1/n)$ estimator using a G estimator, with a suitable selection of α_i parameters. Heuristics as in rows 9 and 10 can give much improvement in some cases at the cost of no improvement at all and even a huge variance in other cases, thus they are not robust heuristics. More sophisticated heuristics as in rows 11 and 12, based on the theoretically justified heuristic in row 8, show a robust behavior.

A very promising estimator is the one in row 7, for all i all μ_i' equal (to μ), in all five examples the variance is very near the optimal $V[F]$. Observe that this estimator has a strong theoretical justification, derived from the study of G estimator. The $\{\alpha_i\}$ parameters that approximate for all i, $\mu_i' = \mu$ could be obtained in an adaptive way; however, a practical way to obtain this estimator should be devised.

Table 3. We show the metric $E_F^{-1} = V[F] \cdot Cost[F]$ and $E_G^{-1} = V[G] \cdot Cost[G]$, i.e., the inverse of efficiency or product of variance and cost for the F estimator and for the optimal G estimator, using for both the same $\{\alpha_k\}$ values. The optimal efficiencies of G estimator, for each set of $\{\alpha_k\}$ values, are computed using the left-hand side of Equation (37), while the efficiencies of F estimator use the right-hand side of Equation (37). We consider for the five numerical examples using an equal count of samples, count inversely proportional to the variances of independent estimators [8,9], and for the three estimators defined in [28]. The sampling costs are (1, 6.24, 3.28). In Example 5, we present the case with equal costs (1, 1), and different costs (1, 5).

	Ex. 1		Ex. 2		Ex. 3		Ex. 4		Ex. 5 costs = (1,1)		Ex. 5 costs = (1,5)	
Estimator	**F**	**G**	**F**	**G**	**F**	**G**	**F**	**G**	**F**	**G**	**F**	**G**
$\alpha_k \propto \frac{1}{n_k}$	102.26	89.40	17.24	15.44	37.47	31.80	98.68	83.78	0.28	0.23	0.83	0.40
$\alpha_k \propto \frac{1}{c_k v_k}$ [8]	49.53	41.29	9.28	8.10	4.03	3.85	300.12	294.85	0.31	0.26	2.76	2.33
$\alpha_k \propto \frac{1}{c_k m_k^2}$ [28]	54.36	46.2	9.82	8.49	3.12	2.49	534.37	449.33	0.20	0.15	1.51	1.00
$\alpha_k \propto \frac{\sigma_{k,eq}}{\sqrt{c_k}}$ [28]	81.43	69.88	13.54	11.67	28.68	23.17	91.01	73.54	1.00	0.98	2.90	2.50
$\alpha_k \propto \frac{M_{k,eq}}{\sqrt{c_k}}$ [28]	79.73	67.77	13.08	11.35	25.74	20.76	31.77	25.90	0.29	0.24	2.72	2.28

Table 4. Variances of Examples 1–5 for some notable cases. The variables $\mu'_{i,1/n}, \sigma'_{i,1/n}$ correspond to $\{\alpha_i = \frac{1}{n}\}$. The variables m_i^2, v_i are the second moment and variance of independent technique i, $v_i = m_i^2 - \mu^2$.

	Estimator	**Example 1**	**Example 2**	**Example 3**	**Example 4**	**Example 5**
1	$Optimal_\alpha \, V[\mathcal{F}]$	22.7122	4.1949	0	0	0.0910
2	$Optimal_\alpha V[F]$	22.7122	4.1944	0	0	0.0903
3	$Optimal_{\alpha,\beta} V[G]$	22.7122	4.1932	0	0	0.0601
4	$V[F(\alpha_i = 1/n)]$	29.1634	4.9175	10.6877	28.1431	0.2771
5	$Optimal_\beta V[G(\alpha_i = 1/n)]$	29.0908	4.9069	10.6125	27.9412	0.2264
6	$Optimal_\alpha V[G(\beta_i = 1/n)]$	27.1603	4.7265	0	0	0.0653
7	$V[F(\alpha_i = arg \, \forall i, \mu'_i \approx \mu)]$	22.8216	4.1980	0	0	0.0926
8	$V[G(\alpha_i = arg \, \forall i, \alpha_i \sigma'_i \approx equal, \beta_i = 1/n)]$	28.4089	4.8313	12.4705	11.2072	0.0734
9	$V[G(\alpha_i \propto \frac{m_i^2}{v_i}, \beta_i = 1/n)]$, [19]	28.0224	4.8513	3.7955	172.192	0.2832
10	$V[G(\alpha_i \propto \frac{1}{v_i}, \beta_i = 1/n)]$	27.4126	4.7897	0.5466	1256.48	0.2943
11	$V[G(\alpha_i \propto \frac{1}{\sigma'_{i,1/n}}, \beta_i = 1/n)]$	28.3977	4.8291	11.95	13.7926	0.0724
12	$V[G(\alpha_i \propto \frac{\mu'_{i,1/n}}{\sigma'_{i,1/n}}, \beta_i = 1/n))]$	27.4426	4.7373	7.8615	6.7895	0.1112
13	$V[G(\alpha_i = arg \, \forall i, \mu'_i \approx \mu, \beta_i = 1/n)]$	41.8791	8.8798	0.0014	0	0.0739

7. Conclusions

In this paper, we have proposed a multiple importance sampling estimator that combines samples simulated from different techniques. The novel estimator generalizes the Monte Carlo balance heuristic estimator, widely used in the literature of signal processing, computational statistics, and computer graphics. In particular, this estimator relaxes the connection between the coefficients that select the number of samples per proposal, and the samples that appear in the mixture of techniques at the denominator of the importance weight. This flexibility shows a relevant improvement in terms of variance in the combined estimator with regard to the balance heuristic estimator (which is included as a particular case in the novel estimator). We have studied the optimal choice of the free coefficients in such a way that the variance of the resulting estimator is minimized. In addition, numerical results have shown that the significant gap in terms of variance between both estimators justifies the use of the novel estimator whenever possible. In the particular, but much used case, of equal sampling, our new estimator shows a big potential for improvement. Future work may include the application of this variance-reduction technique in the context of adaptive importance sampling [15] or within MCMC-based methods that include re-weighting [32].

Appendix A. Difference between the Variances of Deterministic and Randomized Multiple Importance Sampling Estimators

Proof of Equation (9). The difference between the variances of the deterministic multiple importance sampling estimator, Z, and the randomized one, $\mathcal{Z}$, is given by [17] (we normalize here to one sample)

$$
\begin{aligned}
V[\mathcal{Z}^1] - V[Z^1] &= \sum_i \alpha_i \mu'^2_i - \mu^2 \\
&= \sum_i \alpha_i \mu'^2_i - \left(\sum_i \alpha_i \mu'_i \right)^2,
\end{aligned}
\tag{A1}
$$

which is positive, as by Cauchy-Schwartz

$$
\left(\sum_i \alpha_i \mu'_i \right)^2 \leq \left(\sum_i \alpha_i \right) \left(\sum_i \alpha_i \mu'^2_i \right),
\tag{A2}
$$

and equality only happens (apart from the case when both $V[Z^1], V[\mathcal{Z}^1]$ are zero) when for all i, $\alpha_i \propto \alpha_i \mu'^2_i$, i.e., all μ'_i are equal. $\square$

Proof of Equation (10). Observe that if $f(x) \geq 0$ then for all i $\mu'_i \geq 0$, and we have

$$
\left(\sum_i \mu'_i \right)^2 \geq \sum_i \mu'^2_i,
\tag{A3}
$$

and thus

$$
\mu^2 = \left(\sum_i \frac{1}{n} \mu'_i \right)^2 \geq \frac{1}{n} \left(\sum_i \frac{1}{n} \mu'^2_i \right),
\tag{A4}
$$

where equality would be approached when there is an index k such that for all $i \neq k$, $\mu'_k >> \mu'_i$. Thus when $\alpha_i = 1/n$,

$$
V[\mathcal{Z}^1] - V[Z^1] \leq (n-1)\mu^2.
\tag{A5}
$$

$\square$

In general, the more different the μ'_i are, the higher the difference between the variances.

Appendix B. Proof of Theorem 2: Optimal Variance of *F*

Proof. The $\{\alpha_i\}_{i=1}^{n}$ values for the optimal variance of *F* estimator can be obtained using Lagrange multipliers with the target function

$$\Lambda(\{\alpha_i\}_{i=1}^{n}, \lambda) = \sum_{i=1}^{n} \alpha_i {\sigma'}_i^2 + \lambda \left(\sum_{i=1}^{n} \alpha_i = 1 \right).$$

Taking partial derivatives with respect to α_j,

$$
\begin{aligned}
\frac{\partial \Lambda(\{\alpha_i\}_{i=1}^{n}, \lambda)}{\partial \alpha_j} &= \frac{\partial \left(\sum_{i=1}^{n} \alpha_i {\sigma'}_i^2 \right)}{\partial \alpha_j} + \frac{\partial (\lambda (\sum_{i=1}^{n} \alpha_i - 1))}{\partial \alpha_j} \\
&= \sum_{i=1}^{n} \frac{\partial \left(\alpha_i {\sigma'}_i^2 \right)}{\partial \alpha_j} + \lambda = 0.
\end{aligned}
\tag{A6}
$$

The partial derivatives are equal to

$$\frac{\partial \left(\alpha_i {\sigma'}_i^2 \right)}{\partial \alpha_j} = \chi_i(j) {\sigma'}_j^2 + \alpha_i \frac{\partial \left({\sigma'}_i^2 \right)}{\partial \alpha_j}, \tag{A7}$$

where $\chi_i(j)$ is the characteristic function, $\chi_i(j) = 1$ if $i = j$ and 0 if $i \neq j$.

$$
\begin{aligned}
\frac{\partial \left({\sigma'}_i^2 \right)}{\partial \alpha_j} &= \frac{\partial \left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x} - (\mu'_i)^2 \right)}{\partial \alpha_j} \\
&= \frac{\partial \left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x} \right)}{\partial \alpha_j} - 2\mu'_i \frac{\partial \mu'_i}{\partial \alpha_j} \\
&= -2 \int \frac{f^2(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x}))^3} d\mathbf{x} - 2\mu'_i \frac{\partial \mu'_i}{\partial \alpha_j}.
\end{aligned}
\tag{A8}
$$

Since we can write

$$
\begin{aligned}
\frac{\partial \mu'_i}{\partial \alpha_j} &= \frac{\partial \left(\int \frac{f(\mathbf{x}) p_i(\mathbf{x})}{\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x})} d\mathbf{x} \right)}{\partial \alpha_j} \\
&= -\int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x},
\end{aligned}
\tag{A9}
$$

thus Equation (A8) reads

$$
\begin{aligned}
\frac{\partial \left({\sigma'}_i^2 \right)}{\partial \alpha_j} &= -2 \int \frac{f^2(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x}))^3} d\mathbf{x} \\
&\quad + 2\mu'_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^{n} \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x}.
\end{aligned}
\tag{A10}
$$

Then,

$$
\begin{aligned}
\lambda = -\sum_{i=1}^{n} \frac{\partial\left(\alpha_i \sigma'^2_i\right)}{\partial_{\alpha_j}} &= -\sigma'^2_j + 2\sum_{i=1}^{n} \alpha_i \times \\
&\quad \left(\int \frac{f^2(\mathbf{x})p_i(\mathbf{x})p_j(\mathbf{x})}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^3}d\mathbf{x} - \mu'_i \int \frac{f(\mathbf{x})p_i(\mathbf{x})p_j(\mathbf{x})}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^2}d\mathbf{x} \right) \\
&= -\sigma'^2_j + 2\int \frac{f^2(\mathbf{x})p_j(\mathbf{x})(\sum_{i=1}^{n}\alpha_i p_i(\mathbf{x}))}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^3}d\mathbf{x} \\
&\quad - 2\sum_{i=1}^{n}\alpha_i\mu'_i \int \frac{f(\mathbf{x})p_i(\mathbf{x})p_j(\mathbf{x})}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^2}d\mathbf{x} \\
&= -\sigma'^2_j + 2(\sigma'^2_j + \mu'^2_j) - 2\sum_{i=1}^{n}\alpha_i\mu'_i \int \frac{f(\mathbf{x})p_i(\mathbf{x})p_j(\mathbf{x})}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^2}d\mathbf{x} \\
&= \sigma'^2_j + 2\mu'^2_j - 2\sum_{i=1}^{n}\alpha_i\mu'_i \int \frac{f(\mathbf{x})p_i(\mathbf{x})p_j(\mathbf{x})}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^2}d\mathbf{x}
\end{aligned}
\tag{A11}
$$

This is, for all j, the following values have to be equal

$$
\sigma'^2_j + 2\mu'^2_j - 2\sum_{i=1}^{n}\alpha_i\mu'_i \int \frac{f(\mathbf{x})p_i(\mathbf{x})p_j(\mathbf{x})}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^2}d\mathbf{x}
\tag{A12}
$$

Multiplying by α_j and adding over all j in Equation (A11),

$$
\begin{aligned}
\lambda &= \sum_{j=1}^{n}\alpha_j\lambda = \sum_{j=1}^{n}\alpha_j\sigma'^2_j + 2\sum_{j=1}^{n}\alpha_j\mu'^2_j \\
&\quad - 2\sum_{i=1}^{n}\alpha_i\mu'_i \int \frac{f(\mathbf{x})p_i(\mathbf{x})(\sum_{j=1}^{n}\alpha_j p_j(\mathbf{x}))}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^2}d\mathbf{x} \\
&= \sum_{j=1}^{n}\alpha_j\sigma'^2_j + 2\sum_{j=1}^{n}\alpha_j\mu'^2_j - 2\sum_{i=1}^{n}\alpha_i\mu'^2_i \\
&= \sum_{j=1}^{n}\alpha_j\sigma'^2_j
\end{aligned}
\tag{A13}
$$

which is the optimal variance of estimator F. Equation (A12) becomes, for all j

$$
\sum_{j=1}^{n}\alpha_j\sigma'^2_j = \sigma'^2_j + 2\mu'^2_j - 2\sum_{i=1}^{n}\alpha_i\mu'_i \int \frac{f(\mathbf{x})p_i(\mathbf{x})p_j(\mathbf{x})}{(\sum_{k=1}^{n}\alpha_k p_k(\mathbf{x}))^2}d\mathbf{x}.
\tag{A14}
$$

Observe that all derivatives in (A6) are negative for the optimal $\{\alpha_i^\star\}$ values and equal to $-\sum_{j=1}^{n}\alpha_j^\star\sigma'^{\star 2}_j$. □

Appendix C. Derivation of Case 3

Proof of Theorem 6. We have to optimize the target function

$$
\Lambda(\{\alpha_i\}_{i=1}^{n}, \lambda) = \sum_{i=1}^{n} \frac{\alpha_i^2\sigma'^2_i}{\beta_i} + \lambda\left(\sum_{i=1}^{n}\alpha_i - 1\right).
$$

Taking partial derivatives with respect to α_i, as the β_i values are constant,

$$\frac{\partial \Lambda(\{\alpha_i\}_{i=1}^n, \lambda)}{\partial \alpha_j} = \sum_{i=1}^n \frac{1}{\beta_i} \frac{\partial\left(\alpha_i^2 \sigma'^2_i\right)}{\partial \alpha_j} + \lambda = 0. \tag{A15}$$

The partial derivatives are equal to

$$\frac{\partial\left(\alpha_i^2 \sigma'^2_i\right)}{\partial \alpha_j} = 2\alpha_j \chi_i(j)\sigma'^2_j + \alpha_i^2 \frac{\partial\left(\sigma'^2_i\right)}{\partial \alpha_j}, \tag{A16}$$

where $\chi_i(j)$ is the characteristic function. Using the result in Equation (A10), we obtain

$$
\begin{aligned}
\sum_{i=1}^n \frac{1}{\beta_i} \frac{\partial\left(\alpha_i^2 \sigma'^2_i\right)}{\partial \alpha_j} &= 2\frac{\alpha_j \sigma'^2_j}{\beta_j} - 2 \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \\
&\times \left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k p_k(\mathbf{x}))^3} d\mathbf{x} - \mu'_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x} \right) \\
&= -\lambda.
\end{aligned}
\tag{A17}
$$

In Equation (A17), we multiply by α_j, and add over all indexes j, obtaining

$$
\begin{aligned}
\lambda &= -2 \sum_{j=1}^n \frac{\alpha_j^2 \sigma'^2_j}{\beta_j} + 2 \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \\
&\times \left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x})(\sum_{j=1}^n \alpha_j p_j(\mathbf{x}))}{(\sum_{k=1}^n \alpha_k p_k(\mathbf{x}))^3} d\mathbf{x} \right. \\
&\quad - \left. \mu'_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x})(\sum_{j=1}^n \alpha_j p_j(\mathbf{x}))}{(\sum_{k=1}^n \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x} \right) \\
&= -2 \sum_{j=1}^n \frac{\alpha_j^2 \sigma'^2_j}{\beta_j} + 2 \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \\
&\times \left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x})}{(\sum_{k=1}^n \alpha_k p_k(\mathbf{x}))^2} d\mathbf{x} - \mu'_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x})}{(\sum_{k=1}^n \alpha_k p_k(\mathbf{x}))} d\mathbf{x} \right) \\
&= -2 \sum_{j=1}^n \frac{\alpha_j^2 \sigma'^2_j}{\beta_j} + 2 \sum_{i=1}^n \frac{\alpha_i^2 \sigma'^2_i}{\beta_i} = 0.
\end{aligned}
\tag{A18}
$$

We remind that $\sum_{j=1}^n \alpha_j = 1$ which disappears in the left-hand side and the second term of the right-hand side equation.

From Equation (A17), the optimal $\{\alpha_j^\star\}_{j=1}^n$ obey

$$
\begin{aligned}
\frac{\alpha_j^\star \sigma'^2_j}{\beta_j} &= \sum_{i=1}^n \frac{\alpha_i^{\star 2}}{\beta_i} \\
&\times \left(\int \frac{f^2(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k^\star p_k(\mathbf{x}))^3} d\mathbf{x} - \mu'_i \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{(\sum_{k=1}^n \alpha_k^\star p_k(\mathbf{x}))^2} d\mathbf{x} \right).
\end{aligned}
\tag{A19}
$$

$\square$

Proof of Equation (39). For the optimal solution to be, for all i, $\alpha_i^\star = \beta_i$ then

$$
\begin{aligned}
\sigma'^2_j &= \int \frac{f^2(\mathbf{x})\left(\sum_{i=1}^n \alpha_i^\star p_i(\mathbf{x})\right) p_j(\mathbf{x})}{\left(\sum_{k=1}^n \alpha_k^\star p_k(\mathbf{x})\right)^3} d\mathbf{x} \\
&\quad - \sum_{i=1}^n \alpha_i^\star \mu_i' \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{\left(\sum_{k=1}^n \alpha_k^\star p_k(\mathbf{x})\right)^2} d\mathbf{x} \\
&= \sigma'^2_j + \mu'^2_j - \sum_{i=1}^n \alpha_i^\star \mu_i' \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{\left(\sum_{k=1}^n \alpha_k^\star p_k(\mathbf{x})\right)^2} d\mathbf{x},
\end{aligned}
$$

and thus the following equation has to hold for all j

$$
\mu'^2_j = \sum_{i=1}^n \alpha_i^\star \mu_i' \int \frac{f(\mathbf{x}) p_i(\mathbf{x}) p_j(\mathbf{x})}{\left(\sum_{k=1}^n \alpha_k^\star p_k(\mathbf{x})\right)^2} d\mathbf{x}. \tag{A20}
$$

□

Appendix D. An Alternative Perspective Based on the χ^2 Divergence

Let us define $\psi_\alpha = \sum_{k=1}^n \alpha_k p_k(\mathbf{x})$ and $\psi_\beta = \sum_{k=1}^n \beta p_k(\mathbf{x})$. Further, let us define $Z_{\frac{f\beta}{\alpha}} = \psi_\beta[\frac{f(\mathbf{x})}{\psi_{\alpha(\mathbf{x})}}]$. We also recall that the χ^2 divergence between the pdf $\widetilde{f}(\mathbf{x}) = \frac{f(\mathbf{x})}{\mu}$ and ψ, is given by

$$
\chi^2(\widetilde{f}, \psi) = \int \frac{\left(\widetilde{f}(\mathbf{x}) - \psi(\mathbf{x})\right)^2}{\psi(\mathbf{x})} d\mathbf{x}. \tag{A21}
$$

Since in the randomized generalized balance heuristic estimator, $\mathcal{G}$, all samples are simulated i.i.d. from α_β, the variance can be expressed as

$$
V_{\psi_\beta}[\mathcal{G}] = \frac{1}{N^2} \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i^2} \sum_{j=1}^{n_i} V_{\psi_\beta}\left[\frac{f(\mathbf{X}_{i,j})}{\psi_\alpha(\mathbf{X}_{i,j})}\right] \tag{A22}
$$

$$
= \frac{1}{N} \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} V_{\psi_\beta}\left[\frac{f(\mathbf{X})}{\psi_\alpha(\mathbf{X})}\right] \tag{A23}
$$

$$
= \frac{1}{N} \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} E_{\psi_\beta}\left[\left(\frac{f(\mathbf{X})}{\psi_\alpha(\mathbf{X})} - Z_{\frac{f\beta}{\alpha}}\right)^2\right] \tag{A24}
$$

$$
= \frac{1}{N} \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \int \left(\frac{f(\mathbf{x})}{\psi_\alpha(\mathbf{x})} - Z_{\frac{f\beta}{\alpha}}\right)^2 \psi_\beta(\mathbf{x}) d\mathbf{x} \tag{A25}
$$

$$
= \frac{1}{N} \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \int \frac{\left(f(\mathbf{x}) - Z_{\frac{f\beta}{\alpha}} \psi_\alpha(\mathbf{x})\right)^2}{\psi_\alpha(\mathbf{x})} \frac{\psi_\beta(\mathbf{x})}{\psi_\alpha(\mathbf{x})} d\mathbf{x} \tag{A26}
$$

$$
= \frac{1}{N} \sum_{i=1}^n \frac{\alpha_i^2}{\beta_i} \int \frac{\left(\mu\widetilde{f}(\mathbf{x}) - Z_{\frac{f\beta}{\alpha}} \psi_\alpha(\mathbf{x})\right)^2}{\psi_\alpha(\mathbf{x})} \frac{\psi_\beta(\mathbf{x})}{\psi_\alpha(\mathbf{x})} d\mathbf{x}, \tag{A27}
$$

where the integral can be seen as a modified χ^2 divergence with two differences: (a) the normalizing constant $Z_{\frac{f\beta}{\alpha}}$ is modified with regard to μ when $\psi_\alpha \neq \psi_\beta$, and (b) there is the ratio $\frac{\psi_\beta}{\psi_\alpha}$ that appears multiplying (evoking an importance weight between the denominator we have used, ψ_α, and the proposal used to simulate, ψ_β). Note that, when $\psi_\alpha = \psi_\beta$, $Z_{\frac{f\beta}{\alpha}} = \mu$, the ratio of mixtures is equal to one, and $V_{\psi_\beta}[\mathcal{G}] = \frac{1}{N}\mu^2\chi^2(f, \psi_\beta)$.

References

1. Veach, E.; Guibas, L. Optimally Combining Sampling Techniques for Monte Carlo Rendering. In Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 6–11 August 1995; pp. 419–428. [CrossRef]
2. Elvira, V.; Martino, L.; Luengo, D.; Bugallo, M.F. Generalized Multiple Importance Sampling. *Stat. Sci.* **2019**, *34*, 129–155. [CrossRef]
3. Owen, A.; Zhou, Y. Safe and Effective Importance Sampling. *J. Am. Stat. Assoc.* **2000**, *95*, 135–143. [CrossRef]
4. Elvira, V.; Martino, L.; Luengo, D.; Bugallo, M.F. Efficient Multiple Importance Sampling Estimators. *IEEE Signal Process. Lett.* **2015**, *22*, 1757–1761. [CrossRef]
5. Elvira, V.; Martino, L.; Luengo, D.; Bugallo, M.F. Multiple importance sampling with overlapping sets of proposals. In Proceedings of the Statistical Signal Processing Workshop (SSP), Palma de Mallorca, Spain, 26–29 June 2016; pp. 1–5.
6. Elvira, V.; Martino, L.; Luengo, D.; Bugallo, M.F. Heretical Multiple Importance Sampling. *IEEE Signal Process. Lett.* **2016**, *23*, 1474–1478. [CrossRef]
7. Sbert, M.; Havran, V.; Szirmay-Kalos, L. Variance Analysis of Multi-sample and One-sample Multiple Importance Sampling. *Comput. Graph. Forum* **2016**, *35*, 451–460. [CrossRef]
8. Havran, V.; Sbert, M. Optimal Combination of Techniques in Multiple Importance Sampling. In Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI '14, Shenzhen China, 30 November–2 December 2014; ACM: New York, NY, USA, 2014; pp. 141–150. [CrossRef]
9. Sbert, M.; Havran, V. Adaptive multiple importance sampling for general functions. *Vis. Comput.* **2017**, *33*, 1–11. [CrossRef]
10. Sbert, M.; Havran, V.; Szirmay-Kalos, L.; Elvira, V. Multiple importance sampling characterization by weighted mean invariance. *Vis. Comput.* **2018**, *34*, 843–852. [CrossRef]
11. Cappé, O.; Guillin, A.; Marin, J.M.; Robert, C.P. Population Monte Carlo. *J. Comput. Graph. Stat.* **2004**, *13*, 907–929. [CrossRef]
12. Cornuet, J.M.; Marin, J.M.; Mira, A.; Robert, C.P. Adaptive Multiple Importance Sampling. *Scand. J. Stat.* **2012**, *39*, 798–812. [CrossRef]
13. Martino, L.; Elvira, V.; Luengo, D.; Corander, J. Layered adaptive importance sampling. *Stat. Comput.* **2017**, *27*, 599–623. [CrossRef]
14. Elvira, V.; Martino, L.; Luengo, D.; Bugallo, M.F. Improving Population Monte Carlo: Alternative Weighting and Resampling Schemes. *Signal Process.* **2017**, *131*, 77–91. [CrossRef]
15. Bugallo, M.F.; Elvira, V.; Martino, L.; Luengo, D.; Míguez, J.; Djuric, P.M. Adaptive Importance Sampling: The past, the present, and the future. *IEEE Signal Process. Mag.* **2017**, *34*, 60–79. [CrossRef]
16. Kondapaneni, I.; Vevoda, P.; Grittmann, P.; Skřivan, T.; Slusallek, P.; Křivánek, J. Optimal Multiple Importance Sampling. *ACM Trans. Graph.* **2019**, *38*, 37. [CrossRef]
17. Veach, E. Robust Monte Carlo Methods for Light Transport Simulation. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1997.
18. Karlík, O.; Šik, M.; Vévoda, P.; Skřivan, T.; Křivánek, J. MIS Compensation: Optimizing Sampling Techniques in Multiple Importance Sampling. *ACM Trans. Graph.* **2019**, *38*, 151. [CrossRef]
19. Grittmann, P.; Georgiev, I.; Slusallek, P.; Křivánek, J. Variance-Aware Multiple Importance Sampling. *ACM Trans. Graph.* **2019**, *38*, 152. [CrossRef]
20. West, R.; Georgiev, I.; Gruson, A.; Hachisuka, T. Continuous Multiple Importance Sampling. *ACM Trans. Graph.* **2020**, *39*, 136. [CrossRef]
21. Kajiya, J.T. The Rendering Equation. In Proceedings of the Computer Graphics (SIGGRAPH '86 Proceedings), Dallas, TX, USA, 18–22 August 1986; Evans, D.C., Athay, R.J., Eds.; Volume 20, pp. 143–150.
22. Robert, C.P. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 2.
23. Elvira, V.; Martino, L.; Closas, P. Importance Gaussian Quadrature. *IEEE Trans. Signal Process.* **2020**, *69*, 474–488. [CrossRef]
24. Elvira, V.; Chouzenoux, E. Langevin-based strategy for efficient proposal adaptation in population Monte Carlo. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 5077–5081.
25. Rubinstein, R.; Kroese, D. *Simulation and the Monte Carlo Method*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 2008.
26. Sbert, M.; Yoshida, Y. Stochastic Orders on Two-Dimensional Space: Application to Cross Entropy. In Proceedings of the Modeling Decisions for Artificial Intelligence—17th International Conference, MDAI 2020, Sant Cugat, Spain, 2–4 September 2020; Lecture Notes in Computer Science; Torra, V., Narukawa, Y., Nin, J., Agell, N., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; Volume 12256, pp. 28–40.
27. Sbert, M.; Poch, J. A necessary and sufficient condition for the inequality of generalized weighted means. *J. Inequalities Appl.* **2016**, *2016*, 292. [CrossRef]
28. Sbert, M.; Havran, V.; Szirmay-Kalos, L. Multiple importance sampling revisited: Breaking the bounds. *EURASIP J. Adv. Signal Process.* **2018**, *2018*, 15. [CrossRef]
29. Cornebise, J.; Moulines, E.; Olsson, J. Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **2008**, *18*, 461–480. [CrossRef]

30. Míguez, J. On the performance of nonlinear importance samplers and population Monte Carlo schemes. In Proceedings of the 2017 22nd International Conference on Digital Signal Processing (DSP), London, UK, 23–25 August 2017; pp. 1–5.
31. Nielsen, F.; Nock, R. On the chi square and higher-order chi distances for approximating f-divergences. *IEEE Signal Process. Lett.* **2014**, *21*, 10–13. [CrossRef]
32. Geyer, C.J. *Reweighting Monte Carlo Mixtures*; Technical Report; University of Minnesota: Minneapolis, MN, USA, 1991.